Available online at www.sciencedirect.com

ScienceDirect

Molecular Phylogenetics and Evolution 42 (2007) 287–297

MOLECULAR PHYLOGENETICS AND EVOLUTION

www.elsevier.com/locate/ympev

Phylogeny of finescale shiners of the genus *Lythrurus* (Cypriniformes: Cyprinidae) inferred from four mitochondrial genes

Jennifer B. Pramuk [b,*], Michael J. Grose [a], Anna L. Clarke [a], Eli Greenbaum [a], Elisa Bonaccorso [a], Juan Manuel Guayasamin [a], Allan H. Smith-Pardo [a], Brett W. Benz [a], Bethany R. Harris [a], Eric Siegfreid [a], Yana R. Reid [a], Nancy Holcroft-Benson [a], Edward O. Wiley [a]

[a] *Department of Ecology and Evolutionary Biology, Natural History Museum and Biodiversity Research Center, The University of Kansas, Lawrence, 1345 Jayhawk Boulevard, KS 66045-7561, USA*

[b] *Department of Integrative Biology, Brigham Young University, Provo, UT 84602, USA*

Received 27 February 2006; revised 8 June 2006; accepted 11 June 2006
Available online 17 June 2006

Abstract

We infer the phylogenetic relationships of finescale shiners of the genus *Lythrurus*, a group of 11 species of freshwater minnows widely distributed in eastern North America, using DNA sequences from the ND2 (1047 bp), ATPase8 and 6 (823 bp), and ND3 (421 bp) mitochondrial protein-coding genes. The topologies resulting from maximum parsimony, Bayesian, and maximum likelihood tree building methods are broadly congruent, with two distinct clades within the genus: the *L. umbratilis* clade (*L. umbratilis* + *L. lirus* + (*L. fasciolaris* + (*L. ardens*, *L. matutinus*))) and the *L. bellus* clade (*L. fumeus* + *L. snelsoni* + (*L. roseipinnis* + (*L. atrapiculus* + (*L. bellus*, *L. algenotus*)))). Support is weak at the base of several clades, but strongly supported nodes differ significantly from prior investigations. In particular, our results confirm and extend earlier studies recovering two clades within *Lythrurus* corresponding to groups with largely "northern" and "southern" geographic distributions. Several species in this genus are listed in the United States as threatened or of special concern due to habitat degradation or limited geographic ranges. In this study, populations assigned to *L. roseipinnis* show significant genetic divergence suggesting that there is greater genetic diversity within this species than its current taxonomy reflects. A full accounting of the biodiversity of the genus awaits further study.
© 2006 Elsevier Inc. All rights reserved.

Keywords: mtDNA; ND2; ND3; ATPase6; ATPase8; Cyprinidae; Teleostei

1. Introduction

Lythrurus is a clade of 11 species of eastern North American minnows characterized by very small scales on the nape; a dorsal fin origin behind the origin of the pelvic fin; a large, oblique, and terminal mouth; and the development of bright red or yellow fins in reproductive males during breeding condition. The monophyly of the genus is corroborated by three morphological synapomorphies: small scales, especially in the predorsal region; reduced anterodorsal squamation; and enlarged urogenital papillae in breeding females (Mayden, 1989; Snelson, 1972; Wiley and Siegel-Causey, 1994).

Herein, we recognize two clades within the genus, based on earlier morphological studies and on our own results based on mtDNA analysis. The *L. umbratilis* clade consists of five species (*L. umbratilis*, *L. lirus*, *L. fasciolaris*, *L. ardens*, and *L. matutinus*) that are widely distributed from the Mississippi and Ohio valleys, in the southern Great Lakes tributaries and southern Ontario and western New York, southward through to eastern Texas and westward to eastern Kansas and Oklahoma (Snelson and Pflieger, 1975). The *L. bellus* clade consists

* Corresponding author. Fax: +1 801 422 0090.
E-mail address: pramuk@byu.edu (J.B. Pramuk).

1055-7903/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2006.06.008

of six species (*L. fumeus*, *L. snelsoni*, *L. bellus*, *L. alegnotus*, *L. roseipinnis*, and *L. atrapiculus*) with distributions centered in the Mississippi River and northern Gulf Coast. The currently recognized taxonomy of *Lythrurus* largely results from a revision of the *L. roseipinnis* complex presented by Snelson (1972). Boschung and Mayden (2004) review more recent taxonomic changes that have resulted in the 11 species now recognized. A map illustrating generalized geographic distributions of the 11 *Lythrurus* species is presented in Fig. 1.

Phylogenetic relationships of *Lythrurus* to other North American minnows remain unclear. Mayden (1989) hypothesized that *Lythrurus* is the sister of a clade composed of the genera *Cyprinella* and *Luxilus* based on three synapomorphies (large cephalic tubercles, submarginal tubercles on body scales, and a triangular palatine). Coburn and Cavender (1992) placed *Pimephales* and *Opsopeodus* in this clade, with *Luxilus* basal to *Lythrurus*. Molecular analyses corroborate a close relationship between *Pimphales* and *Opospoeodus* (Simons et al., 2003) but not their close relationship to *Lythrurus*, *Cyprinella*, and *Luxilus*.

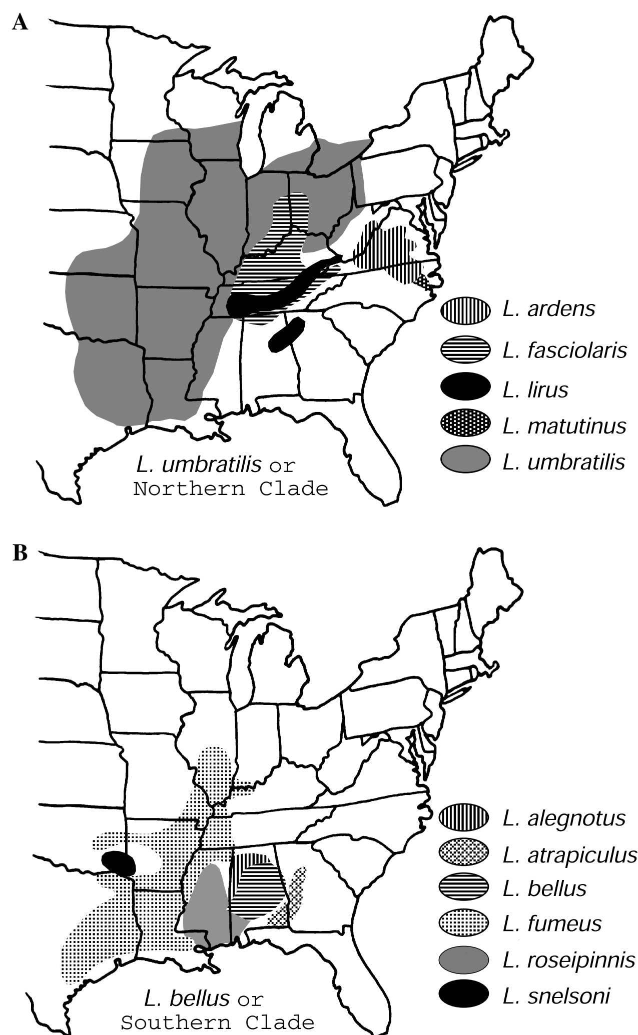

Fig. 1. Map of the generalized geographic distribution of the 11 species of *Lythrurus* in eastern North America corresponding to clades with largely northern (A) and southern (B) distributions. [modified from Schmidt et al. (1998; Fig. 1)].

Until now, no single study of the relationships among all currently recognized species of *Lythrurus* has been accomplished. Stein et al. (1985) used allozymes and distance analysis to study the *L. roseipinnis* group. Mayden (1989) included seven species in his morphological analysis and Wiley and Siegel-Causey (1994) used a combination of allozymes and morphology in a phylogenetic analysis of the same seven species. In their phylogenetic analysis of cytochrome-*b* sequence data, Schmidt et al. (1998) added an additional species (*L. snelsoni*) not analyzed by Wiley and Siegel-Causey. As discussed herein, the only phylogenetic conclusion prior studies appear to agree on is the monophyly of the *L. roseipinnis* species group, composed of *L. alegnotus*, *L. atrapiculus*, *L. bellus*, and *L. roseipinnis*. The primary goal of this study is to test previous hypotheses of relationships among species of *Lythrurus* using the full complement of currently recognized species. In addition, we take the opportunity to revisit the biogeography of the group and to suggest that an understanding of the biodiversity of *Lythrurus* at the species level is still incomplete.

2. Materials and methods

2.1. Taxon sampling

Forty-one specimens representing all 11 species of *Lythrurus* and one specimen from each of two outgroup species, *C. lutrensis* and *Pteronotropis signipinnis*, were sampled. In the absence of strong molecular corroboration for either the "basal" or "apical" position of *Luxilus* (Simons et al., 2003), *C. lutrensis* and *P. signipinnis* were selected as outgroups based on the relationships recovered by Mayden (1989) and Coburn and Cavender (1992). Specimen localities, tissue numbers, and GenBank accession numbers for each sequenced gene region are listed in Table 1.

2.2. DNA extraction, PCR amplification, and sequencing of mtDNA

Total genomic DNA was extracted from small amounts (~25 mg) of frozen or ethanol preserved tissues with a Dneasy Tissue Kit® (Qiagen, Inc.) and visualized on 1% high melt agarose gels in TAE buffer. PCR was performed in 50 μl reactions containing 0.5 U of *Taq* polymerase (Fisher), ~200 ng of genomic DNA, 10 pmol of each primer, 15 nmol of each dNTP, 50 nmol of $MgCl_2$, and buffer (Fisher). Amplification followed published PCR conditions (Palumbi, 1996) and was performed on a BioRad (MyCycler) thermalcycler. Amplified products were purified with AMPure magnetic beads (Agencourt). Cycle sequencing reactions were completed with dye terminator cycle sequencing (DTCS) quick start kit (Beckman Coulter), with minor modifications of the reaction conditions (4 μl kit instead of 8 μl). Sequencing reactions were purified with CleanSEQ magnetic beads (Agencourt) and loaded

Table 1
Taxa of *Lythrurus* and outgroups used in this study, their associated locality data and voucher ID numbers, and GenBank accession numbers for each of the three mitochondrial gene fragments ND2, ATPase8 and 6, and ND3

Taxa	Geographic location	GenBank numbers			
		Voucher ID	ND2	ATPase8 and 6	ND3
Outgroups:					
C. lutrensis	Kansas River, below Bowersock Dam, Lawrence, KS	T530, KU 27255	DQ306610	DQ306653	DQ306567
P. signipinnis	Perdido River, at Florida Route 112 on Alabama border	T5775, KU 33937	DQ306611	DQ306654	DQ306568
Ingroup taxa:					
L. alegnotus 5	Five Mile Creek at old bridge, just downstream of I-59/20 (Valley Creek), AL	UAIC 11055.01-1	DQ306630	DQ306673	DQ306587
L. alegnotus 6	Five Mile Creek at old bridge, just downstream of I-59/20 (Valley Creek), AL	UAIC 11055.01-2	DQ306631	DQ306674	DQ306588
L. alegnotus 7	Five Mile Creek at old bridge, just downstream of I-59/20 (Valley Creek), AL	UAIC 11055.01-3	DQ306632	DQ306675	DQ306589
L. alegnotus 8	Five Mile Creek at old bridge, just downstream of I-59/20 (Valley Creek), AL	UAIC 11055.01-4	DQ306633	DQ306676	DQ306590
L. alegnotus 9	Five Mile Creek at old bridge, just downstream of I-59/20 (Valley Creek), AL	UAIC 11055.01-5	DQ306634	DQ306677	DQ306591
L. ardens 1	Little R. at Dick Binns Rd. ca. 2 mi. SW of Garrett Crossing, VA	UMSL 649.01-1	DQ306640	DQ306683	DQ306597
L. ardens 2	Little R. at Dick Binns Rd. ca. 2 mi. SW of Garrett Crossing, VA	UMSL 649.01-3	DQ306641	DQ306684	DQ306598
L. ardens 3	Little R. at Dick Binns Rd. ca. 2 mi. SW of Garrett Crossing, VA	UMSL 649.01-2	DQ306642	DQ306685	DQ306599
L. ardens 4	Dan River at public access off of NC Hwy 8/89, 2.3 mi N of Hanging Rock, NC	UAIC 12463-1	DQ306643	DQ306686	DQ306600
L. atrapiculus	Panther Creek at US 84, west of Dothan, AL	T5769	DQ306622	DQ306665	DQ306579
L. atrapiculus	Big Horse Creek, FL	T5770	DQ306623	DQ306666	DQ306580
L. atrapiculus	Big Horse Creek at FL 2, 2.5 mi. E of junction of FL 2 and 189	T5771, KU 21896	DQ306624	DQ306667	DQ306581
L. atrapiculus	Boggy Branch of Sepulga River at Highway 31, AL	T5773	DQ306625	DQ306668	DQ306582
L. bellus	Mill Creek at Route 82 just W of Tuscaloosa, AL	T5744	DQ306626	DQ306669	DQ306583
L. bellus	Small Creek, N of Eutaw, 0.1 mi. N of junction of County Road 55 and AL 43 on Ala. Rt. 43, AL	T5745, KU 21871	DQ306627	DQ306670	DQ306584
L. bellus	Oakmulgee Creek, at AL 14, NW of Selma, AL	T5747	DQ306629	DQ306672	DQ306586
L. bellus	Small Creek, N of Eutaw, 0.1 mi. N of junction of County Road 55 and AL 43 on Ala. Rt. 43, AL	T5746, KU 21871	DQ306628	DQ306671	DQ306585
L. fasciolaris 13	Rockcastle River along KY Hwy 89, ca. 1.7 mi NE of KY Hwy 490, KY	UAIC 9851.02-2	DQ306646	DQ306689	DQ306603
L. fasciolaris 14	Rockcastle River along KY Hwy 89, ca. 1.7 mi NE of KY Hwy 490, KY	UAIC 9851.02-1	DQ306647	DQ306690	DQ306604
L. fasciolaris 15	Rockcastle River along KY Hwy 89, ca. 1.7 mi NE of KY Hwy 490, KY	UAIC 9851.02-3	DQ306648	DQ306691	DQ306605
L. fumeus	Lick Creek at TN 144, TN	T5754, KU 23984	DQ306612	DQ306655	DQ306569
L. fumeus	Lick Creek at TN 144, TN	T5755, KU 23984	DQ306613	DQ306656	DQ306570
L. lirus 1	Fourmile Creek at Co. Rd. 49, 1 mi W of Fourmile (Yellowleaf Creek), AL	UAIC 14147.01-1	DQ306614	DQ306657	DQ306571
L. lirus 2	Fourmile Creek at Co. Rd. 49, 1 mi W of Fourmile (Yellowleaf Creek), AL	UAIC 14147.01-2	DQ306615	DQ306658	DQ306572
L. lirus 3	Fourmile Creek at Co. Rd. 49, 1 mi W of Fourmile (Yellowleaf Creek), AL	UAIC 14147.01-3	DQ306616	DQ306659	DQ306573
L. lirus 4	Fourmile Creek at Co. Rd. 49, 1 mi W of Fourmile (Yellowleaf Creek), AL	UAIC 14147.01-4	DQ306617	DQ306660	DQ306574
L. lirus 5	Fourmile Creek at Co. Rd. 49, 1 mi W of Fourmile (Yellowleaf Creek), AL	UAIC 14147.01-5	DQ306618	DQ306661	DQ306575
L. matutinus 10	Tar River at Co. Rd. 1243/1003, 2.4 mi NE of Mitchiners Crossroads, NC	UAIC 11003.03-5	DQ306644	DQ306687	DQ306601
L. matutinus 12	Tar River at Co. Rd. 1243/1003, 2.4 mi NE of Mitchiners Crossroads, NC	UAIC 11003.03-3	DQ306645	DQ306688	DQ306602
L. roseipinnis	Beasha Creek, 4 mi E Laurel Hill on Rt. 488, MS	T5764, KU 22873	DQ306635	DQ306678	DQ306592
L. roseipinnis	Tributary to Brusky Creek, 6.7 miles West of Leakesville on Highway 57, MS	T5767, KU 37856	DQ306638	DQ306681	DQ306595
L. roseipinnis	Small creek flowing into Little River in Claude D Kelly State Park off Highway 21, AL	T5766, KU 37857	DQ306637	DQ306680	DQ306594
L. roseipinnis	Little Creek, 1.5 mi. E of junction of US ? and LA 21 on LA 21, LA	T5765, KU 37854	DQ306636	DQ306679	DQ306593
L. roseipinnis	Little Creek, 1.5 mi. E of junction of US ? and LA 21 on LA 21, LA	T5774, KU 37854	DQ306639	DQ306682	DQ306596
L. snelsoni 16	Mountain Fork River downstream of OK Hwy 4, 0.8 mi SE of Smithville, OK	UAIC 11045.02-4	DQ306649	DQ306692	DQ306606
L. snelsoni 17	Mountain Fork River downstream of OK Hwy 4, 0.8 mi SE of Smithville, OK	UAIC 11045.02-2	DQ306650	DQ306693	DQ306607
L. snelsoni 18	Mountain Fork River downstream of OK Hwy 4, 0.8 mi SE of Smithville, OK	UAIC 11045.02-5	DQ306651	DQ306694	DQ306608
L. snelsoni 19	Mountain Fork River downstream of OK Hwy 4, 0.8 mi SE of Smithville, OK	UAIC 11045.02-1	DQ306652	DQ306695	DQ306609
L. umbratilis	West branch of Mill Creek, 7.6 mi. SW of Alma, KS	T5739	DQ306619	DQ306662	DQ306576
L. umbratilis	West branch of Mill Creek, 7.6 mi. SW of Alma, KS	T5740	DQ306620	DQ306663	DQ306577
L. umbratilis	West branch of Mill Creek, 7.6 mi. SW of Alma, KS	T5741	DQ306621	DQ306664	DQ306578

Table 2
The amplification (a) and sequencing (s) primers used for the continuous ATPase8 and 6 region, ND2, and ND3 regions of the mitochondrial genome along with their respective annealing temperature

Primer Name	Primer sequence (5′ → 3′)	Annealing Temperature (°C)
Lys-L (a, s)	TAG GAA GCT AAA TAT TGG ACA AAG CG	53.4
COIII-H (a, s)	CAT GGG CTT GGA TCA ACT ATA TGA TAG G	53.4
ND2B-L (a, s)	AAG CTT TCG GGC CCA TAC CC	56.7
ND2L2 (s)	ATY CAA ACR GCC CAR GCY TTY GAC C	55
ND2H2 (s)	GGA TTT TAG ATC ATG TGG TTG CAA GGG T	55
ND2E-H (a, s)	TTC TAC TTA AAG CTT TGA AGG C	56.7
Gly-L (a, s)	GTA CAC GTG ACT TCC AAT CA	56.7
Arg-H (a, s)	GGA CTT TAA CCA CAG ACT CTT GAG CCG A	56.7

Primers were developed for this study with the exception of the amplification primers for the ND2 gene (T. Dowling, personal communication).

onto a Beckman Coulter CEQ 8000 XL automated sequencer for detection. The program Sequencher 3.1.1 (Gene Codes Corp.) was used to form contigs and edit sequences. The protein-coding sequences were aligned and translated into amino acids in the application Se-Al v1.0a1 (Rambaut, 1996) to verify alignment. The different gene regions were then concatenated in a single data file for subsequent analyses. The resulting alignments are available on TreeBASE (www.treebase.org/treebase/). Primers and their respective annealing temperatures used for amplification and sequencing are provided in Table 2.

2.3. Sequence analysis

Sequences were examined for base compositional stationarity (i.e., whether base composition is approximately equivalent across all included taxa). The χ^2 statistic generated in the "base composition" menu in PAUP* was used to examine stationarity for each gene as a whole and for third positions of each gene only.

Synonymous codon usage (hereafter referred to as codon bias) was investigated using Wright's (1990) "effective number of codons" (ENC). The ENC can range from 20 (very high codon bias) to 61 (no codon bias). Because of difficulties in calculating ENC for shorter sequences, ENC was calculated for ND2 and ND3 together and for ATPase6 and ATPase8 together. One individual of each species was examined for codon bias except in the case of *L. roseipinnis*, for which two individuals were included to represent different clades within the species.

2.4. Phylogenetic analysis

Each separate gene region, as well all combined data, were analyzed with Modeltest 3.06 PPC (Posada and Crandall, 1998) to find the best model of evolution for the data. Employing the Akaike information criterion (AIC), the model with the lowest AIC score was chosen (Akaike, 1974). Theoretically, the AIC penalizes more complex models and reduces the number of unnecessary parameters that contribute little to describing the data. The most appropriate model of gene evolution for the maximum likelihood (and Bayesian) analysis was estimated for the continuous ATPase8 and 6 fragment and for ND2 and ND3. Modeltest also was run separately on each codon position for coding regions and for the non-coding regions.

Maximum parsimony (MP), Bayesian, and maximum likelihood (ML) analyses were performed on the separate molecular partitions and on the combined data. Initially, data from each of the three DNA fragments (ND2, ATPase6 and 8, and ND3) were analyzed separately. We compared non-parametric bootstrap (npb) values and Bayesian posterior probabilities (bpp) supporting nodes of the resulting trees, and the topologies resulting from the separately analyzed data sets, looking for areas of strongly supported incongruence resulting from two or more data partitions (following Wiens, 1998). Strong support for individual nodes is defined as nodes with bpp $\geqslant .95$ (Alfaro et al., 2003) or npb $\geqslant 70$ (Hillis and Bull, 1993). With the exception of a conflict in the placement of *L. roseipinnis* and *L. fumeus* in the tree derived from the ATPase 6–8 fragment compared to their more "derived" placement in all other analyses, no other strongly supported conflicting relationships were recovered, so all data were combined for all subsequent analyses (Wiens, 1998). We did not employ the incongruence length difference (ILD) test as it has been shown to be a poor test of the compatability of separate data partitions (Hipp et al., 2004).

The data sets were analyzed in combined mixed-model analyses using MrBayes 3.04b (Ronquist and Huelsenbeck, 2003). The analysis of combined data utilized 13 model partitions for the data sets (with each codon position modeled separately for the coding regions). To check for congruence on an identical topology, a minimum of two replicate searches was performed for each separate and combined data set on the BYU BioAg Computational Cluster (http://babeast.byu.edu/index.php). Analyses were initiated with random starting trees and each analysis was run for 20×10^6 generations with four Markov chains employed, the temperature set at 0.4, and with the chain sampled every 1000th generation. The application Tracer (v1.2; Rambaut and Drummond, 2003) was used to view output of the *sump* file generated by MrBayes. Trees generated prior to reaching stationarity were discarded as burn-in. Most analyses reached stationarity relatively quickly (all reached stationarity after 150,000 generations). We then took the resulting 50% majority rule consensus tree as our working hypothesis.

Maximum likelihood (ML) (Felsenstein, 1981) analyses were performed as implemented in PAUP* v4.0b10

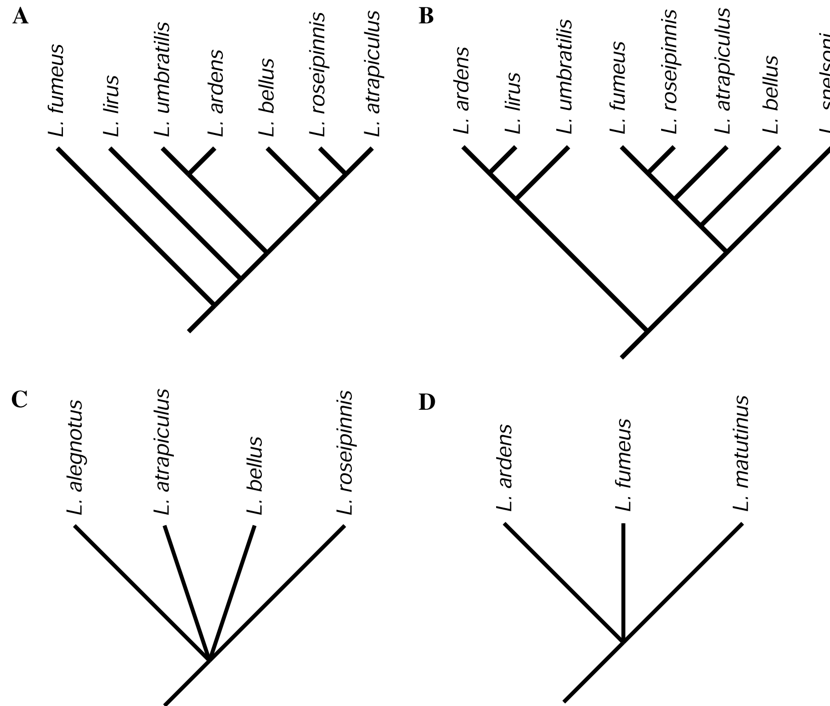

Fig. 2. Maximum parsimony (K–H test) and maximum likelihood (S–H test) optimality criteria were employed to test the illustrated trees in relation to our data and topologies. (A) Parsimony consensus tree presented by Wiley and Siegel-Causey (1994) on the basis of allozyme and morphological data. (B) Phylogenetic tree of Schmidt et al. (1998) derived from ML and MP analysis of cytochrome-*b* data. (C) Hypothesis of the monophyly of the *L. roseipinnis* complex. (D) Hypothesis of the monophyly of the *L. ardens* complex.

(Swofford, 2002). Heuristic searches were performed with 100 random sequence additions and TBR branch swapping. Nodal support was assessed with non-parametric bootstrap (Felsenstein, 1985) with 1000 bootstrap replicates, TBR branch swapping, and 10 random addition replicates. Maximum parsimony (MP) analyses were performed with PAUP* v.4.0b10 using a heuristic search with 10,000 random addition sequence replicates and TBR branch swapping. Nodal support for MP results was assessed through non-parametric bootstrap analysis with 1000 bootstrap pseudoreplicates and one random taxon-addition replicate.

2.5. Hypothesis testing

Alternative phylogenetic hypotheses from the literature were tested under both parsimony and likelihood frameworks. For the parsimony Kishino and Hasegawa (K–H; Kishino and Hasegawa, 1989) test, we employed PAUP* to search for differences in tree scores between all equally optimal trees resulting from constrained searches which were then compared to overall optimal trees. Likelihood topology tests were based on the Shimodaira and Hasegawa (S–H; Shimodaira and Hasegawa, 1999) test as implemented in PAUP*. Ten thousand replicates for every topology test were performed, and the partial likelihoods for each site (RELL model) were resampled. It has been suggested that the S–H test may be conservative in that it may be prone to a certain type of bias so that the number of trees included in the confidence set tends to be very large as the number of trees compared increases (Goldman et al., 2000; Buckley, 2002); however, we agree with these authors in that the S–H test is still safe to employ when the number of possible trees is comparatively small and a relatively large number of characters is included in the analysis. Tree searches were conducted with constraints designed to match tree topologies for

Table 3
Major data partitions analyzed in this study, their characteristics resulting from MP analysis, and the appropriate models selected by Modeltest (Posada and Crandall, 1998) which were employed in separate and combined partitions for Bayesian and ML data analyses

Partition (no. Taxa)	No. characters (pars. inf.)	No. MP trees	TL	CI	RI	ML model
ND2 (43)	1047 (352)	1	1020	0.5412	0.8644	GTR+I +G
ATPase8 (43)	165 (35)	32	89	0.6854	0.8911	GTR with equal rate variation
ATPase6 (43)	659 (158)	15	406	0.6059	0.8824	GTR+G
ND3 (43)	421 (98)	10	238	0.5672	0.8717	GTR with equal rate variation
All combined data (43)	2291 (641)	4	1764	0.5624	0.8683	Mixed model

CI, consistency index (excluding uninformative characters); G, gamma; ML, maximum likelihood; MP, most parsimonious; RI, retention index; TL, tree length.

each hypothesis. Alternative topologies (e.g., constraining monophyly of the *roseipinnis* complex) were constructed in TreeView PPC. The K–H and S–H tests were employed to investigate the following hypotheses: (1) the phylogenetic hypothesis of Wiley and Siegel-Causey (1994) derived from allozyme and morphological data (Fig. 2A); (2) the phylogenetic tree of Schmidt et al. (1998) inferred from analysis of the complete cytochrome-*b* gene (Fig. 2B); (3) monophyly of the *roseipinnis* complex, consisting of *L. alegnotus*, *L. atrapiculus*, *L. bellus*, and *L. roseipinnis* (Fig. 2C); (4) monophyly of the *L ardens* complex (*sensu* Dimmick et al., 1996), consisting of *L. ardens*, *L. fasciolaris*, and *L. matutinus* (Fig. 2D).

3. Results

3.1. Molecules

A total of 2291 bp from four mitochondrial protein-coding genes were obtained: 1047 bp of ND2, 823 bp of the contiguous ATPase8 and 6, and 421 bp of ND3. For ML searches and Bayesian analysis of combined data, a GTR + I + Γ model incorporated the following rate matrix: A-C: 1.0000, A-G: 40.6957, A-T: 1.0000, C-G: 1.0000, C-T: 10.1879, G-T: 1.0000. The shape parameter of the discrete gamma distribution was estimated to be 1.0944 with the proportion of invariant sites equal to 0.5022. Summary statistics of MP analyses for the sepa-

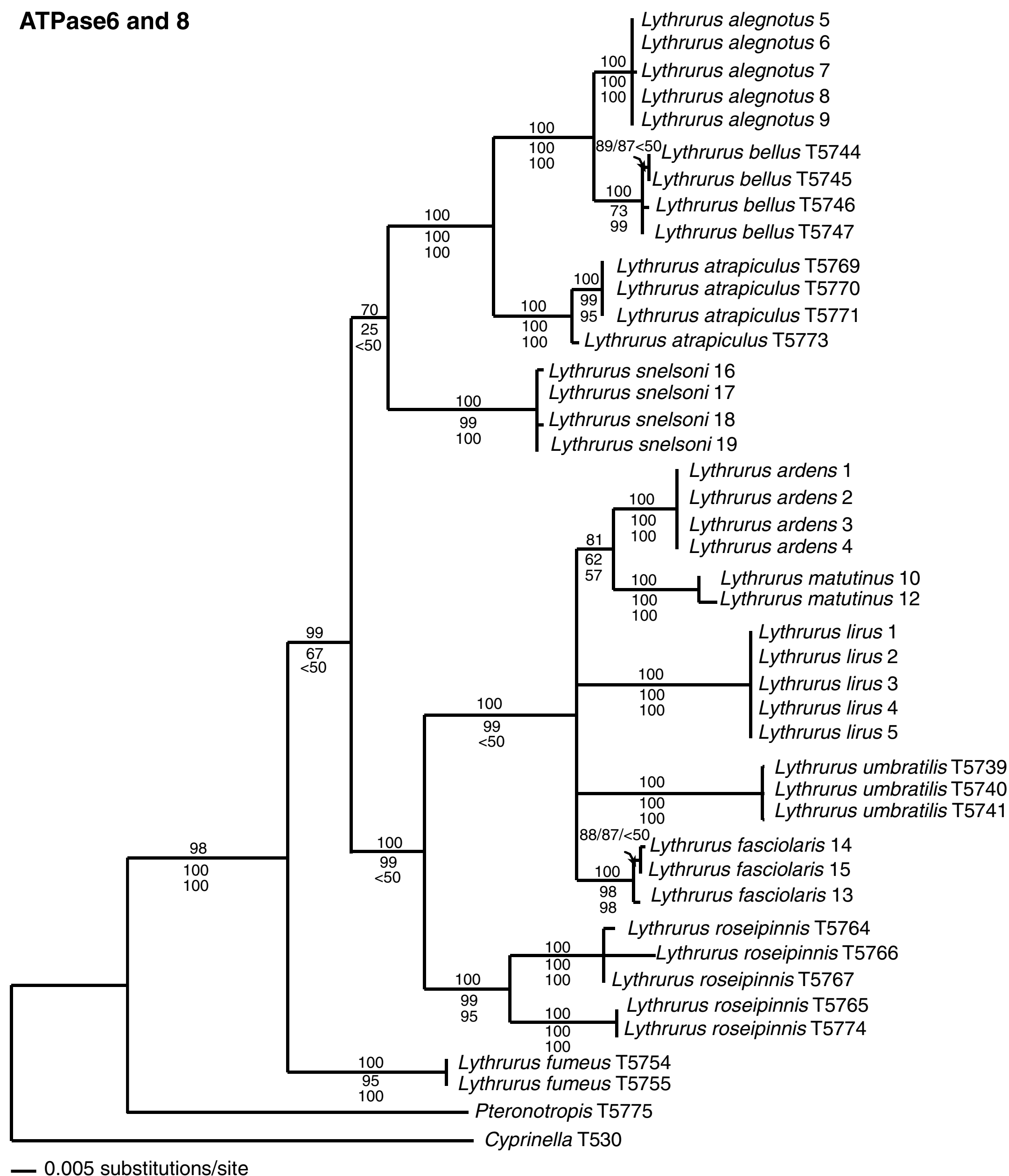

Fig. 3. Bayesian consensus tree resulting from analysis of 824 bp of combined ATPase6 and 8 data, with Bayesian posterior probabilities (multiplied by 100) listed above nodes, and ML bootstrap and MP bootstrap values listed below nodes.

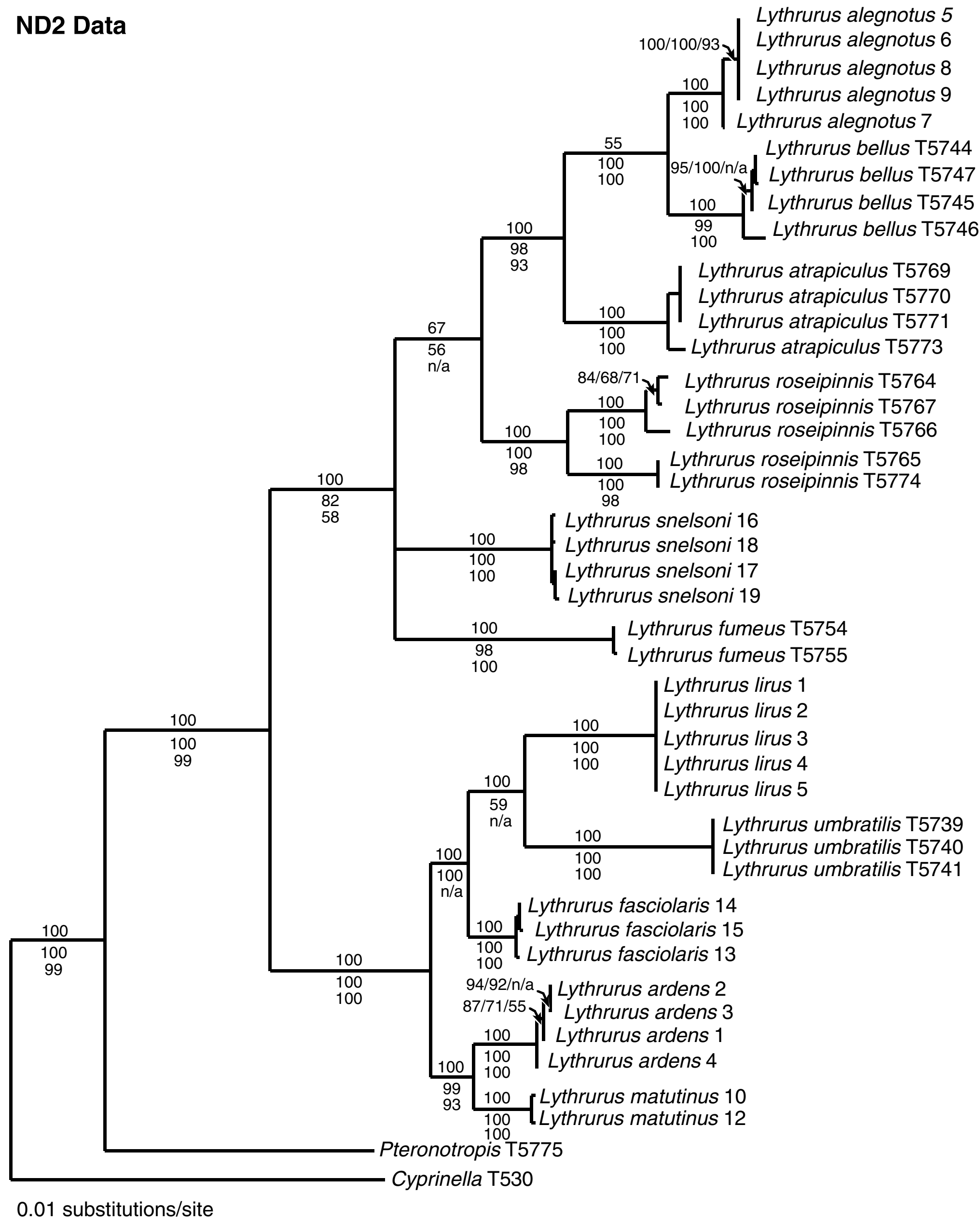

Fig. 4. Bayesian consensus tree resulting from analysis of 1047 bp of ND2 data only, with Bayesian posterior probabilities (multiplied by 100) listed above nodes, and ML bootstrap and MP bootstrap values listed below nodes.

rate and combined data sets and model selected by Modeltest for each gene are shown in Table 3. Base composition across all taxa is stationary for each gene, both as a whole and within third positions only. Codon bias did not appear evident based on the calculated ENC values. For the ND2 + ND3 data set, ENC ranged from 44.8 in *L. snelsoni* to 56.6 in *C. lutrensis,* with all but two individuals having an ENC > 50. For the ATPase6+ATPase8 data set, ENC ranged from 46.9 in *P. signipinnis* to 59.6 in *L. alegnotus*. For this data set, approximately half of the sampled individuals had an ENC > 50. Resulting ENC and χ^2 values are available from the authors upon request.

3.2. Phylogeny of Lythrurus

Maximum parsimony analysis of the combined data resulted in four equally parsimonious trees of 1764 steps (Fig. 3; CI = 0.5624, HI = 0.4376, RI = 0.8683, and RC = 0.4883). Table 3 summarizes the characteristics (e.g., number of parsimony informative characters) for each gene analyzed in this study. Results of model-based methods (Bayesian and ML analyses) are largely congruent with the MP analyses; however, for economy of space, only Bayesian trees resulting from the ND2, ND3, the combined ATPase6 and 8 fragment, and analysis of the combined data are presented and discussed in depth. The ML and MP bootstrap

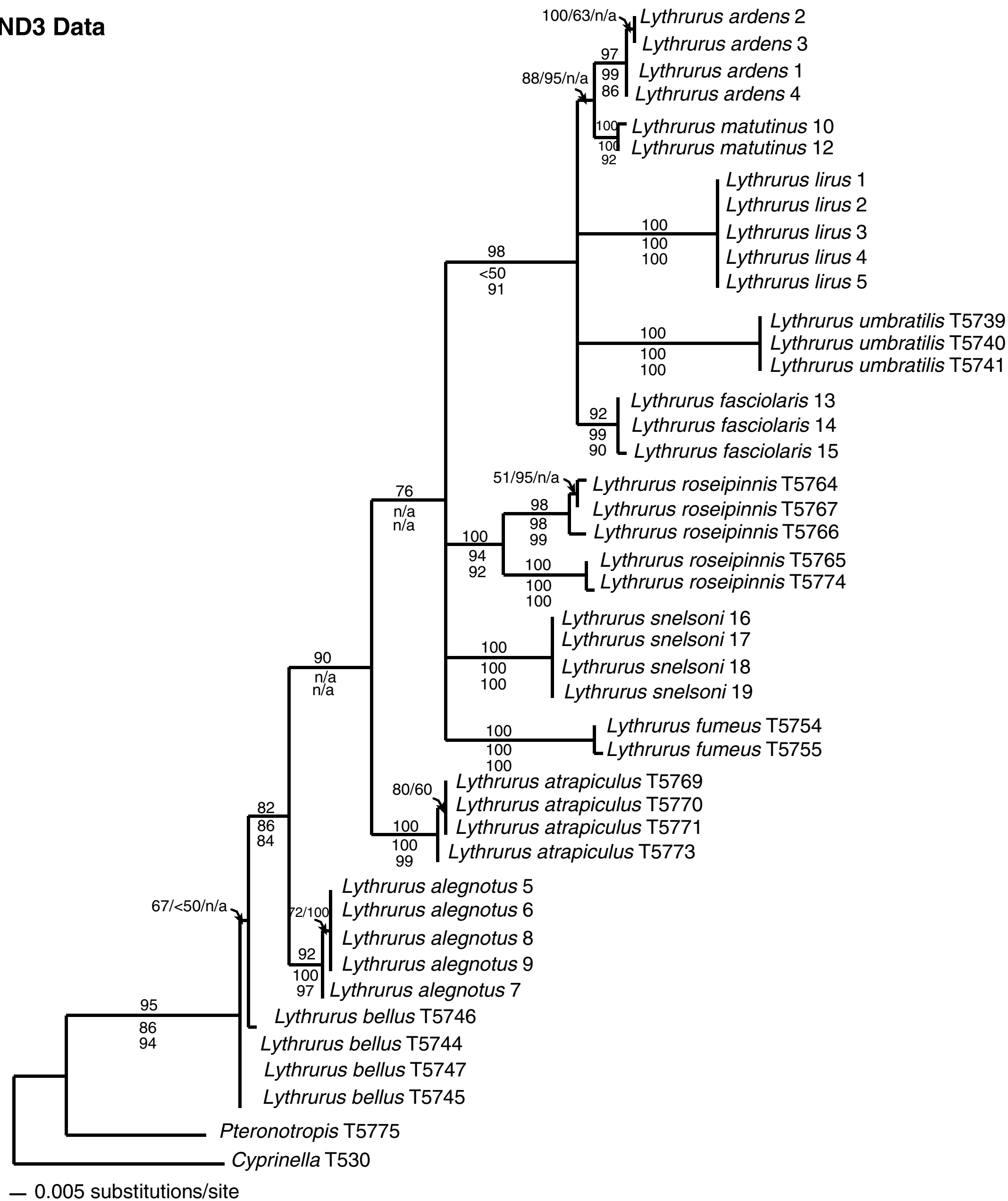

Fig. 5. Bayesian consensus tree resulting from analysis of 421 bp of ND3 data only, with Bayesian posterior probabilities (multiplied by 100) listed above nodes, and ML bootstrap and MP bootstrap values listed below nodes.

values of nodes that are congruent across methods are provided in each of the Bayesian trees (Figs. 3–6). In trees resulting from analyses of most separate genes (except ATPase 6–8) and from the combined data, two major subclades were recovered. The monophyly of the *L. bellus* clade composed of six species (*L. fumeus*, *L. snelsoni*, *L. roseipinnis*, *L. atrapiculus*, *L. bellus*, and *L. alegnotus*) was well supported (bs = 100). Additionally, the *L. umbratilis* clade consisting of five species (*L. umbratilis*, *L. lirus*, *L. fasciolaris*, *L. ardens*, and *L. matutinus*) was well supported (bs = 100). Within the *L. umbratilis* group, *L. umbratilis*, *L. fasciolaris*, and *L. lirus* form a functional polytomy (e.g., bpp < .95) with the well-supported *L. ardens*–*L. matutinus* species pair. Although *L. fasciolaris* grouped with *L. umbratilis* and *L. lirus* rather than its former conspecifics, its placement was not statistically robust. None of the methods or analyses of separate genes or combined data satisfactorily resolved the basal relationships within the *L. bellus* clade. While the combined data support *L. fumeus* as the sister to remaining members of the *L. bellus*, support for this node is not significant in some trees (e.g., results of ND2 data only). In contrast, the tree resulting from the ATPase 6–8 fragment places *L. fumeus* as sister to all other *Lythrurus*. Additionally, the traditionally recognized *L. roseipinnis* group, comprising *L. alegnotus*, *L. atrapiculus*, *L. bellus*, and *L. roseipinnis*, was not recovered with significant support in any analysis. We consider our tree resulting from combined data (i.e.,

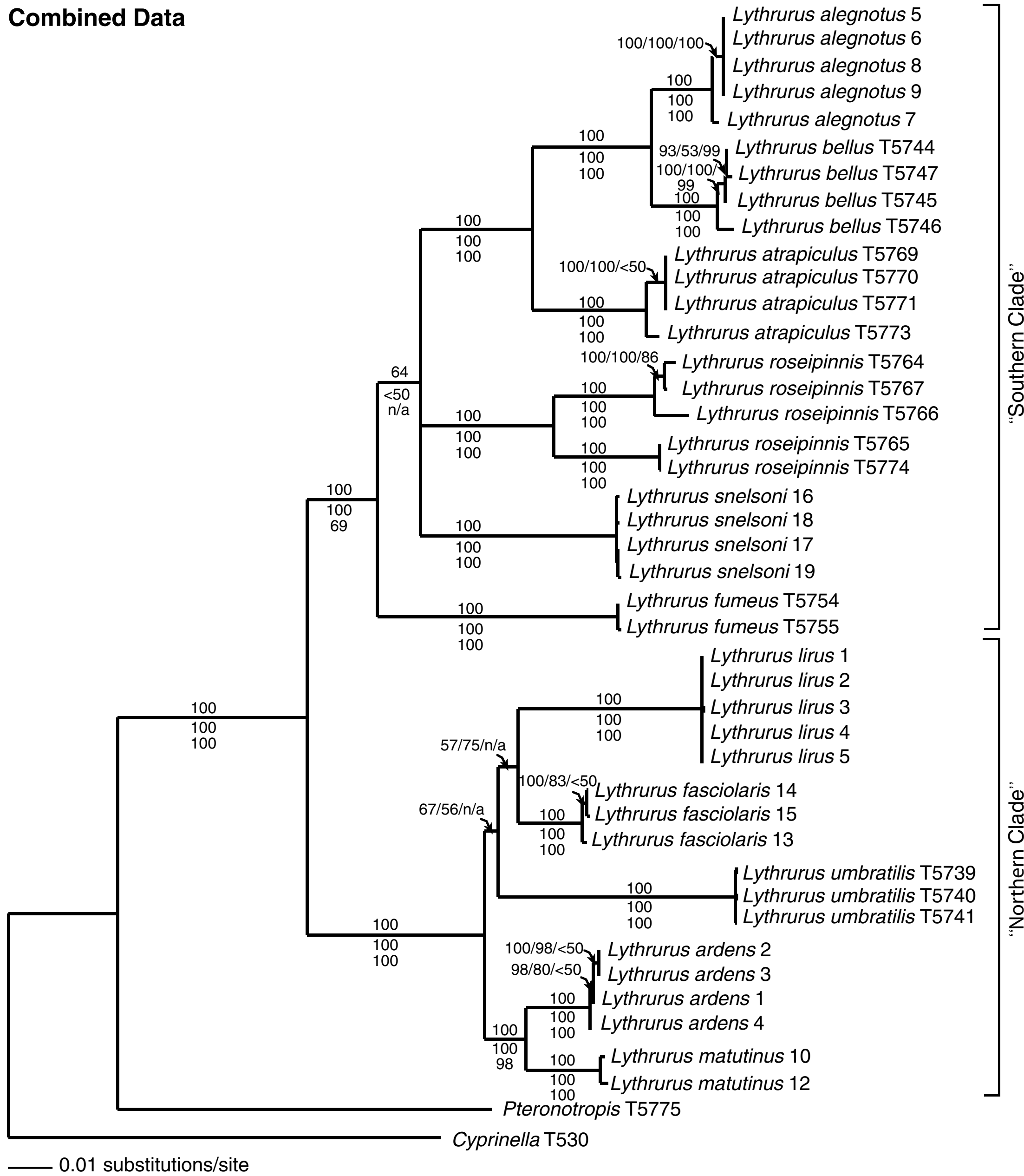

Fig. 6. Bayesian consensus tree resulting from 2291 bp of combined ND2, ATPase8 and 6, and ND3 mitochondrial DNA data. The ML topology was nearly identical to the Bayesian topology. This analysis employed a mixed model, utilizing 13 partitions for each codon position as well as for non-coding regions of the four genes. Bayesian posterior probabilities (multiplied by 100) are listed above nodes and ML bootstrap and MP bootstrap values listed below nodes.

our "total" evidence hypothesis) as our preferred hypothesis of relationships and therefore, focus our following discussion on this tree (Fig. 6).

3.3. Constraint analyses

Using K–H and S–H tests, we tested the topology of Wiley and Siegel-Causey (Fig. 2A) inferred from morphological and allozyme data and the topology of Schmidt et al. (1998) resulting from analysis of cytochrome-*b* data (Fig. 2B) with our data using trees derived from MP and ML analyses. We also tested the hypotheses that the *L. roseipinnis* and *L. ardens* complexes are monophyletic (Fig. 2C and D). The K–H and S–H tests found our topology to be significantly better than the hypotheses of Wiley and Siegel-Causey (1994, Fig.2A)) and Schmidt et al. (1998, Fig.2B)); S–H tests also rejected the

Table 4
Likelihood (S–H) and MP (K–H) topology tests of previous hypotheses of *Lythrurus* relationships as illustrated in Fig. 2A–D

Hypothesis (from Fig. 2)	S–H Δ −ln *L*	*P* value	K–H No. steps (steps shorter)	*P* value
A	255.469	0.0000*	1977 (213)	<0.0001*
B	76.1507	0.0023*	3525 (1730)	<0.0001*
C	47.3921	0.0428*	1764 (0)	1.000
D	32.3038	0.1362	1764 (0)	1.000

Tests indicated with an asterisk are significant at a 0.05 level.

monophyly of the *L. roseipinnis* complex. However, the K–H test was unable to reject the monophyly of either the *L. roseipinnis* or *L. ardens* complexes (Table 4).

4. Discussion

4.1. Evolutionary and biogeographic implications

Our phylogenetic results differ in many respects from previous analyses. The hypothesis that *L. lirus* and *L. fumeus* are either basal (Wiley and Siegel-Causey, 1994) or unresolved (Mayden, 1989) relative to congeners is rejected; each is a member of one of the well-supported subclades.

Within the *L. umbratilis* clade, our results are broadly congruent with Schmidt et al. (1998), given their taxon sampling. They analyzed only one (*L. ardens* s.s. from Virginia) of the three species recognized by Dimmick et al. (1996) that had previously been allocated to *L. ardens* s.l. (Snelson, 1980). While our study does corroborate the monophyly *L. ardens* species trio (*L. ardens*, *L. fasciolaris*, *L. matutinus*), neither Schmidt et al. (1998) nor our study provides strong corroboration for a sister group relationship between *L. lirus* and the trio. Moreover, our study does not find strong support for a relationship between *L. fasciolaris* and the *L. ardens*–*L. matutinus* species pair. Indeed, with regard to the relationship of *L. fasciolaris*, our results are ambiguous: MP weakly supports *L. fasciolaris* with its former conspecifics of the *L. ardens* group, while model-based methods weakly group *L. fasciolaris* with *L. lirus*. Our results do call attention to the distinctive nature of these species in terms of sequence divergence that reflects their distinctive breeding color differences (documented by Dimmick et al., 1996) and to the existence of an Atlantic Drainage clade composed of *L. ardens* and *L. matutinus*.

Our combined data analyses agree with Schmidt et al. (1998) in recovering a monophyletic *L. bellus* clade. However, our data reject their hypothesis that *L. fumeus* is a member of the *L. roseipinnis* species complex (*L. alegnotus*, *L. atrapiculus*, *L. bellus*, and *L. roseipinnis*) referring this species as sister to the *L. roseipinnis* complex + *L. snelsoni*. Moreover, our data do not provide strong corroboration for the monophyly of the *L. roseipinnis* species group, a group long recognized as a distinctive clade (Mayden, 1989; Snelson, 1972; Stein et al., 1985). Interestingly, three morphological synapomorphies corroborate the monophyly of this clade to the exclusion of *L. fumeus*: (1) closely spaced articulation points of the metapterygoid to the hyomandibular and symplectic versus widely spaced; (2) anterodorsal metapterygoid flange gracile versus robust; (3) ventral edge of the posterior flange of the ascending arm of the cleithrum inclined upward versus horizontal (Wiley and Siegel-Causey, 1994). These characters suggest that additional molecular data may recover the group as monophyletic.

Our combined data support *L. atrapiculus* as sister to the *L. bellus*–*L. alegnotus* species pair, refuting the hypothesis of Wiley and Siegel-Causey (1994) stating that *L. atrapiculus* is the sister species of *L. roseipinnis*. Further, *L. roseipinnis* samples in this study showed genetic divergence (4.9–5.3% divergent; uncorrected "*p*" distance) falling within the observed range between currently recognized species (3–12% divergence). Interestingly, Snelson (1972) suggested that *L. roseipinnis* is a complex of at least two distinctive populations. The relatively simple vicariance hypothesis of Wiley and Mayden (1985) accounts for the distribution of recently evolved species of killifishes (Ghedotti and Grose, 1997; Wiley, 1977, 1985), minnows (Grose and Wiley, 2002; Wiley and Titus, 1992), and sand darters (Near et al., 2000; Shaw et al., 1995; Wiley and Hagen, 1997). However, the complex nature of *L. roseipinnis* coupled with vicariance within the Mobile Basin suggests a more complicated series of vicariance events for this clade, as yet unobserved in other groups of fishes (Boschung and Mayden, 2004, review the complexities of biogeographic patterns in the region).

4.2. Conservation issues

Unfortunately, populations of some species of *Lythrurus* are declining in numbers as a result of habitat loss largely related to modern dredging practices and logging. For example, the US Endangered Species Act and/or the IUCN lists *L. alegnotus*, *L. umbratilis*, *L. lirus*, and *L. snelsoni* as threatened, endangered, or of special concern (e.g., Gimenez Dixon, 1996). An accurate assessment of *Lythrurus* biodiversity will aid in making informed conservation decisions. Increasingly, molecular sequence data and DNA barcoding techniques are being used to aid in delimiting cryptic species and have been applied to diverse groups of organisms ranging from vertebrates to insects and plants (e.g., Hebert et al., 2004; Kress et al., 2005; Peppers and Bradley, 2000; Parra-Olea and Wake, 2001). Our results indicate that all currently recognized species of *Lythrurus* are distinct genetically. This study also reports unexpectedly high levels of DNA sequence variation within populations referred to *L. roseipinnis* suggesting that additional, cryptic evolutionary species may yet remain to be described in this lineage. Broader population sampling and additional molecular data including nuclear markers will be required to assess thoroughly the number of evolutionary species present in this geographically widespread clade of fishes.

Acknowledgments

This study was performed as part of *Molecular Sequencing and Systematics*, an upper-level course offered through the Department of Ecology and Evolutionary Biology at KU. Partial support for this work was provided by an IMSD grant from the NIH Minority Opportunities in Research Program (NIGMS; NIH 1 R25 GM 62232). We especially thank Drs. James Orr and Leonard Kristhalka for inviting us to offer this course and EEB and the Biodiversity Research Center and Natural History Museum at KU for obtaining funding to provide necessary equipment and to cover laboratory expenses.

References

Alfaro, M.E., Zoller, S., Lutzoni, F., 2003. Bayes or bootstrap? A simulation study comparing the performance of Bayesian Markov chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence. Mol. Biol. Evol. 20, 255–266.

Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. on Autom. Contr. 19, 716–723.

Boschung, H., Mayden, R., 2004. Fishes of Alabama. Smithsonian Books. Washington D.C., p. 736.

Buckley, T.R., 2002. Model misspecification and probabilistic test of topology: evidence from empirical data sets. Syst. Biol. 51, 509–523.

Coburn, M.M., Cavender, T.M., 1992. Interrelationships of North American fishes. In: Mayden, R.L. (Ed.), Systematics, Historical Ecology, and North American Freshwater Fishes. Stanford University Press, Stanford, CA, pp. 328–373.

Dimmick, W.W., Fiorino, K.L., Burr, B.M., 1996. Reevaluation of the *Lythrurus ardens* (Cypriniformes: Cyprinidae) complex with recognition of three evolutionary species. Copeia 1996, 813–823.

Felsenstein, J., 1981. Evolutionary trees from DNA sequences: A maximum likelihood approach. J. Mol. Evol. 17, 368–376.

Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39, 783–791.

Ghedotti, M.J., Grose, M.J., 1997. Phylogenetic relationships of the *Fundulus nottii* species group (Fundulidae, Cypinodontiformes) as inferred from the cytochrome *b* gene. Copeia 1997, 858–862.

Gimenez Dixon, M., 1996. *Lythrurus snelsoni*. IUCN 2004. 2004 IUCN Red List of Threatened Species (www.iucnredlist.org).

Goldman, N., Anderson, J.P., Rodrigo, A.G., 2000. Likelihood-based tests of topologies in phylogenetics. Syst. Biol. 49, 652–670.

Grose, M.J., Wiley, E.O., 2002. Phylogenetic relationships of the *Hybopsis amblops* species group (Teleostei: Cyprinidae). Copeia 2002, 1092–1097.

Hebert, P.D.N., Penton, E.H., Burns, J.M., Janzen, D.H., Hallwachs, W., 2004. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. Proc. Nat. Acad. Sci. USA 101, 14812–14817.

Hillis, D.M., Bull, J.J., 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst. Biol. 42, 182–192.

Hipp, A.L., Hall, J.C., Sytsma, K.J., 2004. Phylogenetic accuracy, congruence between data partitions, and performance of the ILD. Syst. Biol. 53, 81–89.

Kishino, H., Hasegawa, M., 1989. Evolution of the maximum likelihood estimate of the evolutionary tree topologies from DNA data, and the branching order in Hominoidea. J. Mol. Evol. 29, 170–179.

Kress, J.W., Wurdack, K.J., Zimmer, E.A., Weigt, L.A., Janzen, D.H., 2005. Use of DNA barcodes to identify flowering plants. Proc. Nat. Acad. Sci. USA 102, 8369–8374.

Mayden R.L., 1989. Phylogenetic studies of North American minnows, with emphasis on the genus *Cyprinella* (Teleostei: Cypriniformes). Miscellaneous Publications of the Museum of Natural History, University of Kansas 80, 1–189.

Near, T.J., Porterfield, J.C., Page, L.M., 2000. Evolution of cytochrome *b* and the molecular systematics of *Ammocrypta* (Percidae: Etheostomatinae). Copeia 2000, 701–711.

Palumbi, S.R., 1996. PCR and molecular systematics. In: Hillis, D., Moritz, C., Mable, B. (Eds.), Molecular Systematics, 2nd ed. Sinauer Press, pp. 205–247.

Parra-Olea, G., Wake, D.B., 2001. Extreme morphological and ecological homoplasy in tropical salamanders. Proc. Nat. Acad. Sci. USA 98, 7888–7891.

Peppers, L.L., Bradley, R.D., 2000. Cryptic species in *Sigmodon hispidus*: Evidence from DNA sequences. J. Mammal. 81, 332–343.

Posada, D., Crandall, K.A., 1998. Modeltest: Testing the model of DNA substitution. Bioinformatics 14, 817–818.

Rambaut, A., 1996. Se-Al: Sequence alignment editor. Available from: <http://evolve.zoo.ox.ac.uk/>.

Rambaut, A., Drummond, A., 2003. Tracer v1.2, Available from: <http://evolve.zoo.ox.ac.uk/>.

Ronquist, F., Huelsenbeck, J.P., 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.

Schmidt, T.R., Bielawski, J.P., Gold, J.R., 1998. Molecular phylogenetics and evolution of the cytochrome *b* gene in the cyprinid genus *Lythrurus* (Actinopterygii: Cypriniformes). Copeia 1998, 22–24.

Shaw, K., Wiley, E.O., Titus, T.A., 1995. A phylogenetic analysis of the *Hybopsis amblops* group. Occasional Papers of the Natural History Museum of the University of Kansas 172, 1–27.

Simons, A.M., Berendzen, P.B., Mayden, R.L., 2003. Molecular systematics of North American phoxinin genera (Actinopterygii: Cyprinidae) inferred from mitochondrial 12S and 16S ribosomal RNA sequences. Zoolog. J. Linnean Soc. 139, 63–80.

Snelson Jr., F.F., 1972. Systematics of the subgenus *Lythrurus*, genus *Notropis* (Pisces: Cyprinidae). Bulletin of the Florida State Museum. Biol. Sci. 17, 1–92.

Snelson Jr., F.F., 1980. Systematic review of the cyprinid fish, *Notropis lirus*. Copeia 1980, 323–334.

Snelson Jr, F.F., Pflieger, W.L., 1975. Redescription of the redfin shiner, *Notropis umbratilis*, and its subspecies in the central Mississippi River Basin. Copeia 1975, 231–249.

Stein, D.W., Rogers, J.S., Cashner, R.C., 1985. Biochemical systematics of the*Notropis roseipinnis* complex (Cyprinidae: subgenus *Lythrurus*). Copeia 1985, 154–163.

Swofford, D.L., 2002. Phylogenetic Analysis Using Parsimony (PAUP), version 4.0b10. Computer program distributed by the Illinois Natural History Survey. Champaign, Illinois.

Wiens, J.J., 1998. Combining data sets with different phylogenetic histories. Syst. Biol. 47, 568–581.

Wiley, E.O., 1977. The phylogeny and systematics of the *Fundulus nottii* species group (Teleostei: Cyprinodontidae). Occasional Papers of the Natural History Museum of the University of Kansas 67, 1–31.

Wiley, E.O., Hagen, R.H., 1997. Mitochondrial DNA sequence variation among the sand darters (Percidae: Teleostei). In: Kocher, T., Stepien, C. (Eds.), Molecular Evolution of Fishes. Academic Press, New York, pp. 75–96.

Wiley, E.O., Mayden, R.L., 1985. Species and speciation in phylogenetic systematics, with examples from the North American fish fauna. Ann. Missouri Bot. Gardens 72, 596–635.

Wiley, E.O., Siegel-Causey, D., 1994. A phylogenetic analysis of the *Lythrurus roseipinnis* species complex (Teleostei: Cyprinidae) with comments on the relationships of other *Lythrurus*. Occasional Papers of the Natural History Museum of the University of Kansas 171, 1–20.

Wiley, E.O., Titus, T.A., 1992. Phylogenetic relationships among members of the *Hybopsis dorsalis* species group (Teleostei: Cyprinidae). Occasional Papers of the Natural History Museum of the University of Kansas 219, 1–18.

Wright, F., 1990. The 'effective number of codons' used in a gene. Gene 87, 23–29.